

DOE Genomic Science Program 2014 Strategic Plan

Mission-Driven Systems Biology



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research

DOE Genomic Science Program

Contacts

Joseph Graber

Team Lead: Genomic Science Program
Biological Systems Science Division
Office of Biological and Environmental Research
U.S. Department of Energy Office of Science
301.903.1239

Todd Anderson

Director
Biological Systems Science Division
Office of Biological and Environmental Research
U.S. Department of Energy Office of Science
301.903.3213

genomicscience.energy.gov

DOE Genomic Science Program: Mission-Driven Systems Biology

2014 Strategic Plan

October 2014



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research

Table of Contents

Executive Summary	v
Genomic Science Program: Conceptual Approach and Mission Drivers.....	1
Genomic Science for DOE Missions.....	5
Systems Biology for Bioenergy	5
Systems Biology–Enabled Biosystems Design.....	7
Systems Biology for Sustainable Bioenergy.....	9
Systems Biology to Advance Understanding of Carbon Cycling and Other Biogeochemical Processes....	12
Cross-Cutting Computational Biology, Bioinformatics, and Predictive Modeling	14
Programmatic Context for DOE Genomic Science.....	19
DOE Office of Science	19
DOE Office of Biological and Environmental Research.....	19
Genomic Science Program Management Principles	19
For More Information	Inside back cover
Definitions of Terms.....	Inside back cover
Websites.....	Inside back cover

Executive Summary

Operating within the Department of Energy's (DOE) Office of Science, the Biological and Environmental Research (BER) program supports fundamental research to address critical DOE mission challenges. The broad goals of these research activities are to:

- Expand foundational knowledge of functional properties of biological systems and their underlying design principles to enable the development of secure and sustainable bioenergy solutions.
- Advance understanding of Earth's integrated biogeochemical processes to improve predictive climate models and identify key factors controlling the storage, release, or transport of carbon, nutrients, or contaminants in the environment.

DOE-sponsored research has been critical to developing systems biology into the expansive conceptual approach described in a 2009 report by the National Academy of Sciences (see sidebar, Applying the Power of Systems Biology, this page). High-throughput genome sequencing of microbes, plants, and complex environmental assemblages of organisms has provided the vital blueprint necessary to understand the functional potential of organisms and interactive communities. By examining the translation of genetic codes into integrated networks of regulatory elements, catalytic proteins, and metabolic networks that define all living organisms, systems biology research sheds light on the fundamental principles that govern functional properties of organisms and how their processes respond to community interactions and environmental variables.

DOE's Genomic Science program supports systems biology research aimed at identifying these foundational principles driving biological systems of plants, microbes, and multispecies communities relevant to DOE missions in energy and the environment. Building on the foundation of sequenced genomes and metagenomes, the program focuses on a tightly coupled approach that combines experimental physiology, omics-driven analytical techniques, and computational modeling of functional biological networks. The

program's ultimate objectives (see Fig. 1. Genomic Science Program Goal and Objectives, p. vi) are to:

- Determine the molecular mechanisms, regulatory elements, and integrated networks needed to understand genome-scale functional properties of biological systems.
- Develop omics experimental capabilities and enabling technologies needed to achieve dynamic, systems-level understanding of organism and community function.
- Flexibly scale understanding of biological processes from defined subsystems to individual organisms, consortial assemblies of multiple organisms, or complex communities operating at ecosystem scales.
- Understand the foundational rules and "design principles" governing living systems and develop tools for more sophisticated biosystems design, enabling the targeted modification of functional properties at the genome scale.
- Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance predictive understanding and manipulation of biological systems.

Advancing fundamental knowledge of these systems bridges critical knowledge gaps that must be addressed to enable biological solutions to crucial DOE missions. Development



Applying the Power of Systems Biology

The ongoing "omics" revolution continues to generate new and increasingly sophisticated experimental systems biology approaches and analytical techniques that can be brought to bear on Department of Energy mission goals. A useful description of systems biology is provided in the National Academy of Sciences report, *A New Biology for the 21st Century*:

"Systems biology seeks a deep quantitative understanding of complex biological processes through dynamic interaction of components that may include multiple molecular, cellular, organismal, population, community, and ecosystem functions. It builds on foundational large-scale cataloguing efforts (e.g., genomics, proteomics, metabolomics, etc.) that specify the 'parts list' needed for constructing models. The models relate the properties of parts to the dynamic operation of the systems they participate in."

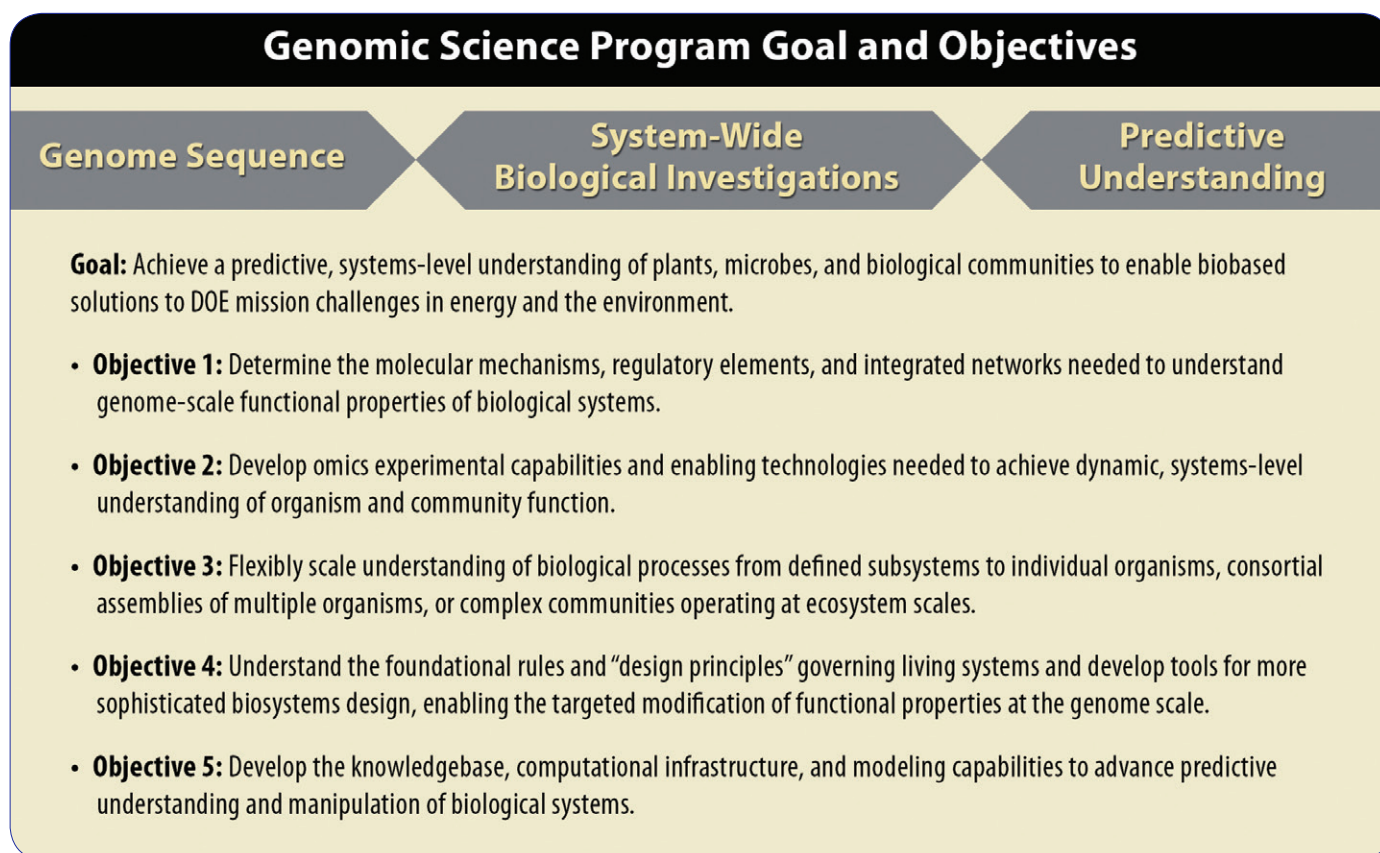


Fig. 1. Genomic Science Program Goal and Objectives.

of innovative approaches for sustainable bioenergy production will be accelerated by systems biology–based understanding of nonfood plants that can serve as dedicated bioenergy feedstocks and microbes capable of deconstructing biomass into their sugar subunits and synthesizing next-generation biofuels (either from cellulosic biomass or through direct photosynthetic capture of carbon dioxide). Genomic Science program research also employs the omics-driven tools of modern systems biology to analyze interactions between organisms that form biological communities and their surrounding environments. Understanding the

relationships between molecular-scale functional biology and ecosystem-scale environmental processes illuminates the basic mechanisms that drive biogeochemical cycling of metals and nutrients, carbon biosequestration, and greenhouse gas emissions in both terrestrial ecosystems impacted by climate change and in agricultural systems producing bioenergy feedstocks. As new understanding emerges, strategic emphasis will be placed on overcoming knowledge gaps and fostering transformational breakthroughs that most effectively address DOE mission-critical research objectives in energy and the environment.

Genomic Science Program: Conceptual Approach and Mission Drivers

Genomes of individual organisms and metagenomes (i.e., the total genomic information of biological communities) contain the information and operating capabilities that determine structure and function across multiple scales of biological organization, from molecular to ecosystem. Exploring biological systems across all scales in a comprehensive and integrated way is essential to understanding how these systems operate in nature or in more applied contexts related to new technology endpoints for DOE missions (see Fig. 2. Multiscale Explorations for Systems Understanding, this page). Systems biology emphasizes the holistic, multidisciplinary study of complex interactions that specify the function of an entire biological system—whether single cells, multicellular organisms, or multiorganism communities—rather than the reductionist study of individual components in isolation.

For biological systems central to DOE missions in energy and the environment, Genomic Science program research analyzes properties and processes on three fundamental levels.

- **Molecular:**

Focusing on genes, proteins, macromolecular complexes, and other biomolecules that provide structure and perform a cellular function. Such an approach aids in understanding how the genome determines dynamic biological structure and function at all scales, from genes to ecosystems, and how proteins and protein

complexes function individually or in interactions with other cellular components.

- **Cellular:** Investigating dynamic molecular processes, networks, and subsystems controlled and coordinated to enable complex cellular processes such as growth and metabolism.
- **Multicellular and Multiorganismal Systems:** Exploring diverse cellular systems that interact to carry out coordinated complex processes that both respond to and alter their environments to determine how cells work in the context of multicellular tissues of plants and multiorganism communities of microbes and plants.

The myriad biological structures and processes that exist within these three system levels are interconnected and coordinated by an intricate set of regulatory controls and continuous interactions with the surrounding environment. To investigate biology at multiple scales, the Genomic Science

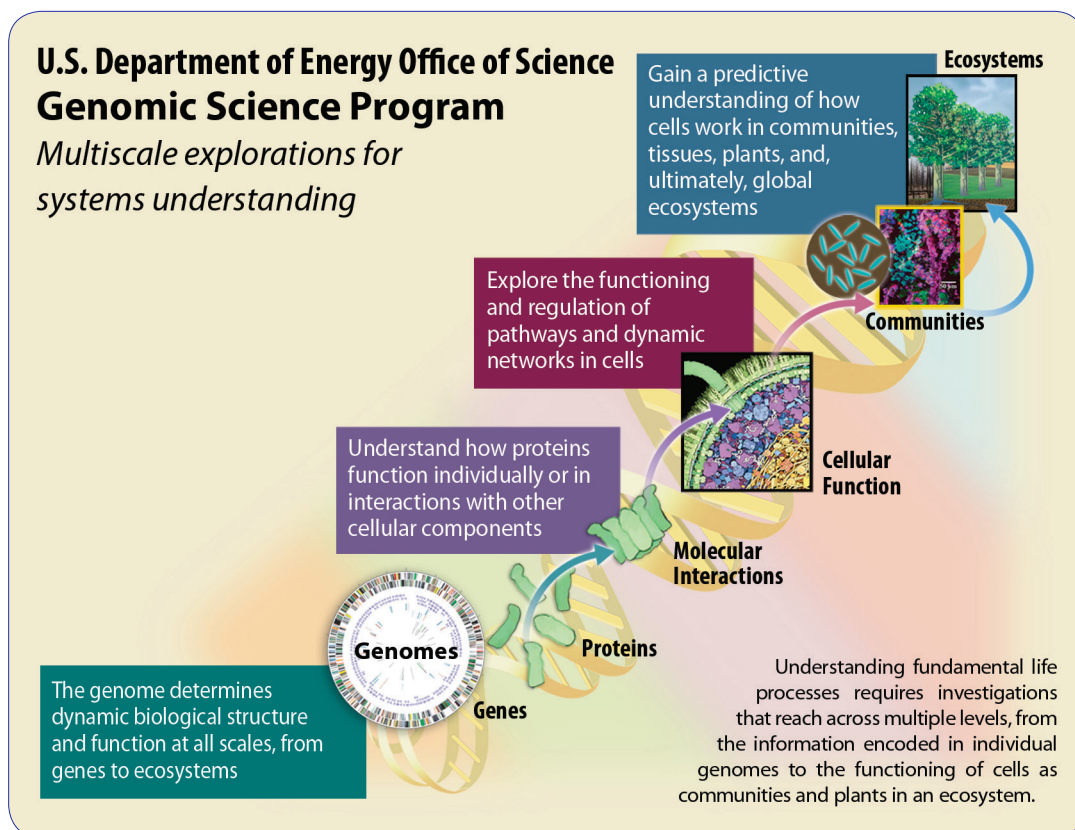


Fig. 2. Multiscale Explorations for Systems Understanding.

program is building new multidisciplinary research communities and advancing development of next-generation, automated technologies that increase sample throughput and analytical reliability while reducing analysis time. Key research technology and methodology development areas for the Genomic Science program include genomics, analytical omics, molecular imaging and structural analysis, modeling and simulation, and genome-scale engineering tools that span all three levels of organization (see sidebar, Systems Biology Tools, Techniques, and Technologies, this page).

As a leader in systems biology research, the Genomic Science program builds on a foundation of genome sequences, genetic regulatory networks, densely arrayed metabolic pathways, and higher-scale organismal interactions to identify the common fundamental principles that drive living systems. Knowledge of these common principles revealed by studying organisms relevant to any one DOE mission facilitates breakthroughs in basic biology important to other DOE and national needs. By leveraging the increasing availability of sequences from whole organism genomes and

Systems Biology Tools, Techniques, and Technologies

Genomics and Metagenomics. Sequencing and analyzing DNA from individual organisms (genomics) or microbial communities in environmental samples (metagenomics) form the foundation of systems biology research. The sequences of Department of Energy (DOE)-relevant microbes, plants, and communities are provided largely by the DOE Joint Genome Institute (JGI), an important DOE user facility and a world leader in generating microbial, metagenome, plant, and other genome sequences (see sidebar, BER National User Facilities, p. 20). In addition to sequencing and annotating genomes, Genomic Science program research seeks to develop improved methods for rapidly identifying and characterizing functional and regulatory gene networks in microbes, microbial communities, and plants.

Analytical Omics. Transcriptomics, proteomics, and metabolomics—collectively described as omics analyses—identify and measure the abundance and fluxes of key molecular species indicative of organismal or community activity. Global analyses of RNA transcripts, proteins, and metabolites inform scientists about organisms' physiological status. Such analyses, as well as stable isotope tracking and nano secondary ion mass spectrometry (NanoSIMS) characterization techniques, also provide insights into gene function and indicate which genes are activated and translated into functional proteins as organisms and communities develop or respond to environmental cues. Metamethods that analyze DNA, RNA, and proteins extracted directly from environmental communities enable discovery of new biological processes and provide novel insights into relationships between the composition of communities and functional processes that they perform.

Molecular Imaging and Structural Analysis. Genomic Science program investigators are developing and using new methods for characterizing the chemical reaction surfaces, organization, and structural components in molecular complexes and tracking molecules to view cellular processes as they are occurring. Depending on the spatial scale, a variety of imaging technologies can be used to visualize the complex molecular choreography within biological systems. Some of these structural and imaging tools (e.g., synchrotrons, neutron sources, and electron microscopes) are available or are being developed at DOE Office of Science national user facilities. These facilities provide photon, neutron, electron, magnetic, and mechanical instrumentation with state-of-the-art spatial, temporal, and chemical sensitivity.

Predictive Modeling. Computational models are used to capture, integrate, and represent current knowledge of biology at various scales. For example, researchers are using genome sequence and other comprehensive datasets (molecular, spatial, and temporal data) to build models of signaling networks, gene regulatory circuits, and metabolic pathways.

Genome-Scale Engineering. Growing interest in biological design research stems from its wide-ranging potential. Knowledge of the principles that govern biology will enable predictions of biological system behavior under changing conditions. Moreover, tailoring the behavior of these systems for defined purposes will necessitate their reengineering or the design of new systems. Novel genome-scale engineering tools and technologies, in turn, will improve understanding of natural systems and their response to changing environmental conditions.

environmental samples (metagenomes), Genomic Science program researchers are developing advanced methods to facilitate the translation of genome sequence into predictive understanding of function. These methods cut across DOE missions in energy and the environment.

Genomic Science program systems biology research extends traditional scientific methodology by addressing complex problems through coordinated research among interdisciplinary teams with complementary expertise in biological, physical, and computational sciences. A nested array of experimental techniques and analytical technologies must be created, refined, and deployed to investigate and understand biological systems, linking different levels of biological discovery to gain a predictive understanding of whole systems—from cells to ecosystems. To accomplish the desired in-depth understanding of biological systems, an unprecedented integration of experimental biology, analytical technologies, data and computation, theory, modeling,

simulation, and experimentation must occur. For many program research priorities, a team-oriented approach emphasizing collaboration, communication, and research integration is often critical to the success of the overall program, and this approach is reflected in the diversity of mechanisms used to support Genomic Science program research. These mechanisms range from small collaborative teams of researchers to large-scale, multidisciplinary research centers involving dozens of scientists.

DOE has a successful history of conducting this scale of scientific research and developing the tools and concepts needed to address high-risk, complex problems underlying achievement of DOE mission goals. In this tradition, the Genomic Science program research enterprise (initiated in 2002; formerly called the Genomes to Life program and subsequently Genomics:GTL) is pursuing solutions to biological grand challenges critical to addressing DOE mission goals.

Genomic Science for DOE Missions

Systems Biology for Bioenergy

Rationale

The vast majority of liquid transportation fuel used in the United States currently is derived from petroleum. In addition to the significant problems associated with security and renewability of these resources, their continued use results in massive releases of carbon dioxide and other greenhouse gases that drive global climate change. In recent years, the expanded availability of ethanol from corn starch and of biodiesel from soybeans has supplemented the transportation fuel supply but raised concerns regarding competition between biofuel production and the food supply. Also, given the energy-intensive agricultural practices normally used to produce these crops, the degree to which a shift to these types of biofuels would actually reduce overall greenhouse gas emissions has been the subject of much debate.

Cellulosic plant biomass (i.e., fibrous or woody plant materials such as stems and leaves) can be broken down into its component sugars by a combination of physical, chemical, and enzymatic treatments. These sugars can then be converted into ethanol or other liquid biofuels by fermentative microbes or other chemical processes. Biofuels derived in this manner have the potential to provide a secure, renewable source of energy that will reduce dependence on fossil fuels and emissions of greenhouse gases. Agricultural residues (i.e., nonedible parts of crop plants) represent one current source of biomass carbon available for fuel production, but supply of this material is unlikely to meet total demands. Every year, a significant percentage of biomass must be left to decay in fields to maintain soil stock composition; furthermore, food crops are limited in their ability to grow on marginal lands, which comprise most of the landscape available for bioenergy production.

Dedicated biomass crops (i.e., switchgrass, poplar, energy cane, and a variety of other plant species) offer one possible route toward sustainable biofuels production that does not directly compete with food crops and could be grown on lands unsuitable for food agriculture. Dedicated biomass feedstocks potentially could yield much greater amounts of cellulosic material and have decreased water and fertilizer requirements and greater tolerance to pests and disease.

In addition to using dedicated biomass plants to produce sugars suitable for biofuels production, photosynthetic microbes such as algae and cyanobacteria potentially offer direct capture of atmospheric carbon dioxide and conversion to biofuels in a single step.

Despite these possible advantages, biofuels produced from cellulosic biomass or via direct microbial photosynthesis are not yet widely available. Compared with the relatively simple sugar chain of starch, cellulosic biomass is a complex, heterogeneous material that is much more difficult to degrade into its component sugars and currently requires expensive chemical pretreatments and enzyme cocktails. This recalcitrance of cellulosic biomass to degradation is due to the crystalline structure of cellulose in plant cell walls, as well as to the presence of covalently bound lignin polymers, which make lignocellulose a very stable material. As such, most cellulosic biofuels are not yet cost competitive with either fossil fuels or ethanol produced from corn starch. Ethanol, whether produced from corn or cellulosic materials, has considerably lower energy density than gasoline or diesel fuel and is more difficult to transport via current distribution infrastructures or use as the primary fuel constituent in most existing vehicle engines. Attempts at commercially viable biofuels production by photosynthetic microbes have been hampered by the significant front-end costs required to establish and maintain scalable microbial cultivation systems and difficulties in engineering production of biofuel compounds by these organisms.

A variety of significant research and development barriers need to be addressed before cellulosic biofuels can be more broadly adopted for use. Technical hurdles particularly relevant to the core strengths of the Genomic Science program include:

- Limited understanding of plant cell wall structural properties that impart strength and resistance to degradation, hindering the development of better biomass deconstruction strategies.
- Inefficiency and high cost associated with currently available physical, chemical, and enzymatic treatments for breakdown of cellulosic biomass and conversion of the resulting sugars to biofuel compounds.

- Gaps in fundamental understanding of systems biology properties of (1) plant species with the potential to serve as biomass feedstock crops and (2) microorganisms capable of mediating deconstruction of complex plant biomass, photosynthetic capture of atmospheric carbon dioxide, and synthesis of advanced biofuel compounds.
- Limitations in currently available approaches for genome-scale engineering and targeted biodesign aimed at improving biomass feedstock crop yields or microbial conversion of biomass and synthesis of advanced biofuels compatible with existing distribution pipelines and engines.
- Elucidation of the regulation of gene networks, proteins, and metabolites to improve plant feedstock productivity and sustainability and to advance understanding of carbon partitioning and nutrient cycling.
- Comparative approaches to enhance knowledge of the structure, function, and organization of plant genomes, leading to innovative strategies for feedstock characterization, breeding, manipulation, and improvement.
- Characterization of plant germplasm collections and advanced breeding lines of bioenergy crops to discover and deploy valuable alleles for key bioenergy traits.

Systems Biology Approaches

DOE's Genomic Science program is uniquely well suited to address the fundamental research challenges embedded within these technical barriers (see Fig. 3. From Biomass to Biofuels, this page), and the integrative systems biology approach pioneered by the program lends itself to the multidisciplinary nature of the problems. For example, significant advances in breeding, molecular genetics, and genomic technologies provide new opportunities to (1) build on existing knowledge of plant biology and (2) more confidently predict and manipulate functional properties of potential biomass feedstock plant species. Applying a systems biology approach permits high-throughput characterization of the genes, proteins, and molecular interactions that influence cellulosic biomass production and enables accelerated improvement of biomass feedstocks for enhanced yields and growth properties. Specific targets include:

- Development of new cultivars of regionally adapted bioenergy feedstock crops in public breeding programs using innovative approaches to identify desirable traits and accelerate trait integration.
- Research into the complex interactions between bioenergy feedstock plants and their environment and how these processes influence plant growth and development, expression of bioenergy-relevant traits, and adaptation to changing environments.

Similarly, continuing advances in genome-enabled systems biology approaches for microorganisms relevant to biofuels production provide a number of opportunities to address key barriers. Only a limited number of microbes that are capable of complex biomass breakdown, photosynthetic capture of carbon dioxide, or synthesis of potential biofuels have been studied sufficiently to be considered “model organisms” that can be reliably manipulated for

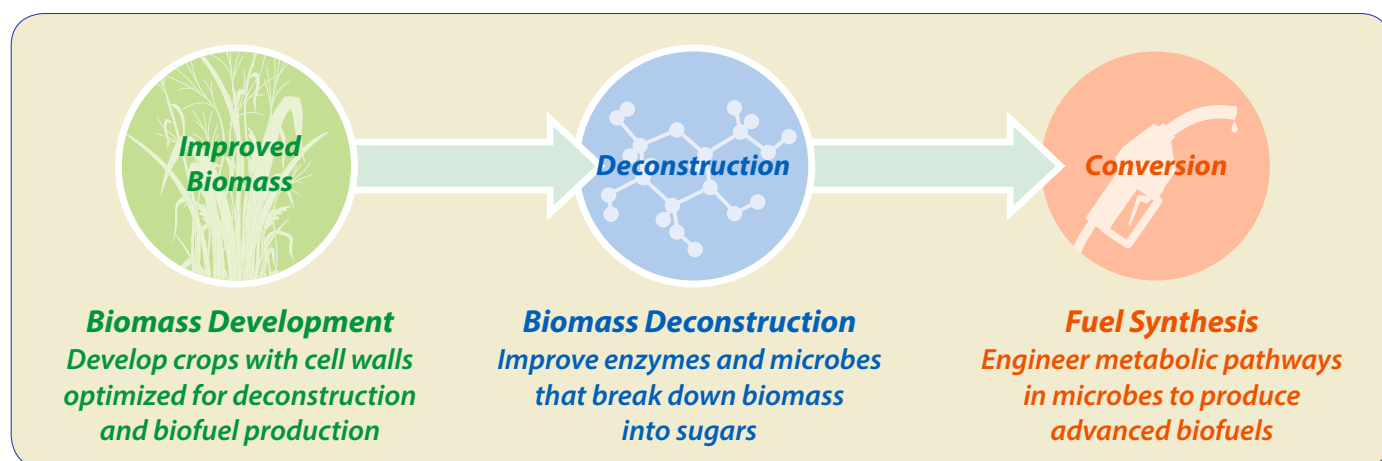


Fig. 3. From Biomass to Biofuels.

experiments. The subset of those with potential to serve as “chassis organisms” for industrial biotechnology applications is even smaller. Systems biology research on microbes with these capabilities will not only advance our knowledge of the fundamental nature of these functional processes but also facilitate engineering of microbes having these traits:

- Increased rates of biomass degradation and expanded utilization of biomass components (i.e., hemicellulose and lignin monomers).
- Decreased susceptibility to product inhibition by toxins released during biomass breakdown or high titers of synthesized biofuels, thus increasing the efficiency and yield of biofuel production.
- Increased tolerance to physicochemical stresses associated with industrial-scale biofuels production (i.e., elevated temperature, altered pH levels, or fuel toxicity).
- Modified functional properties that result in increased yield of biofuel compounds or synthesis of a broader range of molecules that can be used as next-generation biofuels and related bioproducts.

In addition, coupling these results with systems biology work on biomass-degrading microbes potentially could enable development of new consolidated biomass-processing approaches, in which a single organism (or assemblage of organisms) performs coupled deconstruction of biomass and synthesis of biofuel compounds. DOE’s Bioenergy Research Centers are spearheading these research efforts (see sidebar, DOE Bioenergy Research Centers, p. 8).

More detailed information about Genomic Science program objectives relating to biofuels production can be found at genomicscience.energy.gov/biofuels/.

Systems Biology–Enabled Biosystems Design

Rationale

Multidisciplinary systems biology approaches are necessary for analyzing disparate types of genome-scale data to enable a deeper understanding of biological systems embedded within both plants and microbes. Recent advances in the integration of computational biology with omics-enabled analytical technologies allow complex biological networks to be dynamically modeled and displayed. Genomics and systems-level predictive

understanding of biological systems, pioneered by the Genomic Science program, are uncovering foundational design rules that govern system behavior to the extent that rational genome-scale redesign of organisms is becoming possible. Also referred to as “synthetic biology,” this new but rapidly developing field already is providing novel tools for collaborating teams of biologists and chemical engineers to construct biological systems and organisms that address unique challenges and enhance understanding of complex biological systems.

The merging of biology, chemistry, physics, and engineering has the potential to transform fundamental and applied science by shedding light on the basic principles of biological system organization and evolution. This understanding can then be used to extend and enhance the capabilities of natural organisms to solve significant practical problems associated with the production of biofuels and related coproducts from renewable biomass. Extensive modification of existing networks or design of specific synthetic systems can both advance the understanding of biological systems and provide novel tools for interrogating basic biological function. Just as early trial and error in airplane design refined the understanding of fundamental fluid dynamics laws and enabled today’s fully automated designs, iterative design cycles in biology will reveal the complex interconnections among biological laws, experimental observations of biological systems, and design of new biological behaviors.

A number of recent breakthroughs have piqued keen interest in genome engineering and biological design. New regulatory circuits can be constructed and evaluated for high-level function and control, and complete metabolic pathways can be assembled, engineered, and introduced into living cells to produce high-value compounds. Entire bacterial genomes can be replaced with modified synthetic counterparts. Orthogonal molecular processes have been developed to incorporate unnatural amino acids into proteins, conferring new functions by codon replacements through directed evolution. Novel nucleases with customizable specificity for genome engineering have been developed from bacterial transcription activator–like effectors (TALEs) and the type II clustered, regularly interspaced, short palindromic repeats (CRISPR)/Cas9 system. Substantial progress also has been made toward constructing synthetic eukaryotic chromosomes. Finally, the vast amount of comprehensive data available for genes,

transcripts, proteins, and metabolites under different conditions for multiple individual organisms has dramatically advanced network analysis and computational modeling of biological systems.

Systems Biology Approaches

These advances open new doorways to understand the foundational principles governing the systems properties of living organisms and to develop novel approaches for



DOE Bioenergy Research Centers

To focus the most advanced biotechnology-based resources on the challenges of biofuel production, the Department of Energy (DOE) established three Bioenergy Research Centers (BRCs) in September 2007, with a second 5-year phase beginning in 2012. Informed by the Genomic Science program conceptual approach and operating as flagship multidisciplinary research centers within the program, each BRC represents an integrative partnership with expertise spanning the physical, chemical, biological, and computational sciences, including genomics, microbial and plant biology, analytical chemistry, computational biology and bioinformatics, and engineering. The BRCs are structured to facilitate knowledge sharing among multiple disciplines so that breakthroughs in one area can be capitalized on and translated to other areas of emphasis. In these integrated and collaborative environments, the BRCs pursue the necessary fundamental research to improve the processes needed for large-scale, cost-effective production of advanced biofuels from cellulosic biomass. Additionally, as the centers approach biofuel production challenges from different angles, the types of knowledge gained are multiplied, new questions explored, and new avenues of research pursued, ultimately accelerating the pathway to improving and scaling up biofuel production processes.

The three centers, which were selected through a scientific peer review process, are based in geographically diverse locations—the Southeast, Midwest, and West Coast—with partners across the nation. DOE's Oak Ridge National Laboratory leads the BioEnergy Science Center (BESC) in Oak Ridge, Tennessee; the University of Wisconsin–Madison leads the Great Lakes Bioenergy Research Center (GLBRC) in Madison in partnership with Michigan State University; and DOE's Lawrence Berkeley National Laboratory leads the Joint BioEnergy Institute (JBEI) in Emeryville, California. Partners include a number of universities, private companies, nonprofit organizations, and DOE national laboratories.

Advances resulting from high-risk, high-return BRC research are providing the knowledge needed to accelerate

scientific discovery and develop new bio-based products, methods, and tools crucial to an emerging biofuels industry. For example, through the development of new, high-throughput technologies, the DOE BRCs have created new plant strains optimized for bioenergy production; using advanced genomic techniques, they have discovered numerous new enzymes for deconstructing lignocelluloses into sugars; and, with some of the most advanced approaches to metabolic engineering, the BRCs have re-engineered microbes to produce not just ethanol, but drop-in hydrocarbon substitutes for gasoline, diesel, and potentially jet fuel.

The ultimate goal for the three DOE BRCs is to provide the fundamental science to underpin a cost-effective, advanced cellulosic biofuels industry. Using systems biology approaches, the BRCs are focusing on new strategies to reduce the impact of key cost-driving processes in the overall production of cellulosic biofuels from biomass. For these biofuels to be adopted on a large scale, they must represent environmentally sustainable and economically competitive alternatives to existing fuel systems. New strategies and findings emanating from the centers' fundamental research are addressing three grand challenges for cost-effective advanced biofuels production:

- Develop next-generation bioenergy crops by unraveling the biology of plant development.
- Discover and design enzymes and microbes with novel biomass-degrading capabilities.
- Develop transformational microbe-mediated strategies for advanced biofuels production.

The science needed to solve these complex challenges requires multiple, coordinated, multidisciplinary teams approaching problems from varied perspectives to accelerate scientific progress; this approach is exemplified by the collaborative activities undertaken by each of the BRCs.

More details on the DOE BRCs and their current research activities can be found at genomicscience.energy.gov/centers/.

large-scale manipulation of functional properties that would not be possible via more traditional metabolic engineering approaches. For example, biodesign approaches applied to biomass feedstock plants could lead to new crops that express their own nitrogen-fixing enzymes without the need for bacterial symbionts or that incorporate components of the C4 and crassulacean acid metabolism (CAM) carbon fixation pathway into C3 plants. Such modifications could result in plants with significantly decreased fertilizer requirements, more efficient water utilization, and greatly improved sustainability characteristics. Understanding and improving stress tolerance in microbes is another critical issue that could benefit from biosystems design approaches. Currently, almost every fermentation process is limited by the microbe's tolerance to the final product. A novel way to overcome this problem might be the complete redesign of cell membrane composition by introducing genes to synthesize new membrane lipids. For example, the lipids making up archaeal membranes are very different from those of bacteria. Thus, moving the pathways for lipid biosynthesis from archaea to other microbes could create engineered organisms more tolerant to alcohol and thus more resilient to the stress of biofuel production. In eukaryotes, subcellular compartmentalization in organelles allows cells to contain metabolic pathways and sequester toxic compounds. Engineering cells with repurposed organelles could enable the production of high intracellular concentrations of biofuel molecules without affecting cytoplasm conditions.

Specific knowledge gaps and developmental opportunities in biosystems design that the Genomic Science program can address include:

- Biological design principles to understand foundational laws governing living systems and better define the “solution space” available to biosystems designers.
- Genome-scale engineering tools enabling improved integration of large synthetic DNA constructs, high-throughput gene editing and recoding technologies, and incorporation of alternate amino acids and other synthetic design assemblies.
- New DNA synthesis and assembly technologies for cost-effective, high-efficiency synthesis and assembly of large DNA constructs.
- Diverse platform microbes and plants with sufficient baseline understanding of systems biology properties

and available molecular genetics tools that induce amenability to genome-scale engineering, providing a broader array of functional capabilities and growth characteristics.

- Minimal cell and *in vitro* systems using highly streamlined platform organisms or entirely cell free *in vitro* systems with fully defined pathways and functional components that facilitate easier modification and increased process control.
- Advanced computationally aided design tools for biological systems (e.g., BioCAD) to facilitate *in silico* design of synthetic parts and systems, accommodate a wider suite of biological parts and platform organisms, and enable efficient workflows for construction and experimental validation of large numbers of genetic design variants.
- Biocontainment mechanisms permitting incorporation of multiple redundant safeguards at each step of the design process, thereby preventing inadvertent release or deliberate misuse of synthetic biological systems.

Advancing biosystems design principles in a range of microbial and plant species facilitates exploration of biological solutions toward DOE bioenergy and environmental missions and advances rigorous hypothesis testing of fundamental biological mechanisms.

More detailed information on Genomic Science program objectives in the area of biosystems design can be found at genomicscience.energy.gov/biosystemsdesign/.

Systems Biology for Sustainable Bioenergy

Rationale

Properly designed biofuel feedstock systems can be considerably more sustainable than their grain-based counterparts, avoiding competition with food production and potentially delivering ecosystem services not currently provided by existing systems. Realizing this potential is not necessarily straightforward, however. Feedstocks grown on “marginal” lands (i.e., land that is less fertile and more water stressed and prone to erosion) will be exposed to multiple stressors simultaneously (see Fig. 4. Challenges of Marginal Soils, p. 10). To achieve reliable and sustainably high yields, bioenergy feedstocks must have the capacity to adapt and maintain productivity even in such challenging environments. In particular, feedstocks need an enhanced capacity to use

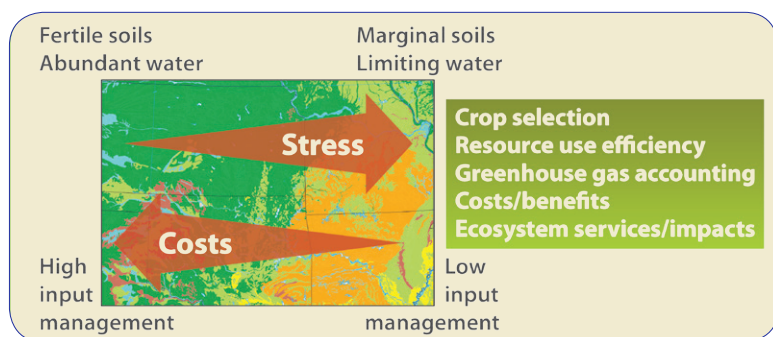


Fig. 4. Challenges of Marginal Soils. Biofuel crops planted on marginal soils with less inherent fertility face stresses to minimize management inputs that would otherwise compromise their climate and other social benefits. These stresses are exacerbated by environmental and economic imperatives. [Soil map background courtesy U.S. Department of Agriculture]

water and nutrients efficiently, acquire nitrogen and phosphorus from nutrient-depleted soils, and withstand pests and disease. These components interact in ways that cannot always be predicted from individual factors, yet it is the net effect of all interactions that lends sustainability attributes. Providing climate mitigation and improved air, soil, and water quality requires knowledge of all the organisms and environmental factors that contribute to the sustainability of highly productive ecosystems.

Recent advances in the systems biology of both plants and microbes have the potential to contribute significantly to the understanding of these interacting pieces and therefore to the design of sustainable biofuel systems. Linking these advances to those in ecosystem science provides an unprecedented opportunity to substantially advance both the fundamental knowledge of systems biology in general and the more specific ability to design sustainable biofuel systems.

Plant productivity, defined as plants' ability to fix carbon dioxide into biomass, is an essential characteristic of sustainable biofuel systems. In most agricultural ecosystems, plant productivity is limited by a deficiency in resources such as light, water, and nutrients. In particular, plant traits enabling the efficient use of water, nitrogen, and phosphorus are likely to become increasingly important as climate change brings about shifting and potentially volatile precipitation and temperature patterns. Resource-efficient, sustainable agroecosystems are capable of converting valuable inputs such as water and nutrients into valuable outputs with minimal waste.

Systems Biology Approaches

To sustainably produce biofuel crops for marginal areas, new breeding strategies are needed for maximizing biomass yields under low-input conditions. In addition to the substantial genetic variation that exists within and between plant species for resource use efficiency traits, a growing body of evidence indicates that the microbial consortia within and surrounding plant roots have a significant influence on the acquisition and utilization of these valuable resources. Genomics approaches can be used to harness these attributes for developing more resource efficient crops by providing a basis for trait-based selection of superior genotypes for breeding programs, particularly in perennial grasses

and trees that have substantially longer breeding cycles. Advances in DNA sequencing technology will facilitate development of novel strategies such as genome-wide association studies (GWAS) to accelerate breeding of superior genotypes.

Microorganisms have a dramatic impact on plant biology, playing a critical role in overall plant fitness, including nitrogen fixation and the scavenging of phosphorus and other critical nutrients from soil. Rather than studying plants as isolated entities, investigating their breeding and agronomic management as part of an integral association with the surrounding microbiota will benefit both the sustainable productivity of bioenergy crops and the ecosystem services associated with large-scale bioenergy cropping systems. This research effort will require an understanding of the genomic and molecular interactions in feedstock plants' immediate microbiome, as well as the biogeochemical processes mediated by microbial communities in surrounding agricultural soils. With the advent of high-throughput sequencing, the feasibility of relating co-occurrence of feedstock cultivars and associated microbiota enables the identification of species and genotype specificity between plant and microbe and reveals the molecular underpinnings of biotic interactions and the community. Enhancing understanding of the genetic rules governing community composition and development could facilitate selection of optimal genotypes for long-term deployment in managed settings.

Bioenergy ecosystems capture and sequester carbon, mitigate greenhouse gas fluxes, regulate water and nutrient flows to

aquatic and other parts of the landscape, and provide habitat for organisms that benefit both crop and natural communities. All these attributes and processes will be affected by the establishment of biofuel crops on lands that now host ecosystems with different plant communities and that are managed with different levels of intensity. This change will result in the delivery of a different set of ecosystem services than before conversion. The net contribution of biofuel crops to environmental sustainability depends on many interacting factors, almost all of which are influenced by how the crop and its associated microbiome interact. These interactions will differ by crop, location, and management; therefore, a fundamental understanding of their effects at the ecosystem scale is necessary to enable predictions of aggregate effects at landscape and regional scales.

Predicting the impact of different bioenergy cropping systems on belowground carbon capture and stabilization, as well as on biogenic production of greenhouse gases, will require an improved understanding of both complex plant-soil-environment interactions at multiple scales and microbial sources of greenhouse gases, including genomic and environmental factors that regulate these phenomena. Subtle variations in plant traits such as root architecture and chemistry, exudation rates, and mycorrhizal associations have the potential to affect these processes. Research is needed to better understand the mechanisms and processes controlling the types and rates of carbon inputs to and outputs from the belowground systems of bioenergy crops; research also must account for variable effects associated with different bioenergy crops, soil types, edaphic conditions, management practices, and climatic regions.

The biological processes underpinning sustainable agricultural systems are inherently complex and have important emergent properties across spatial and temporal scales. Multiscale modeling, which integrates mechanistic models describing system performance at discrete biological scales to evaluate whole system behavior, is an irreplaceable tool for understanding the behavior of complex biological systems and for enabling evaluation of system behavior in a range of contexts, including future climate and management scenarios.

The growing availability of large datasets at both genomic and ecosystem scales and the increasing accessibility and power of computational resources present unprecedented opportunities to develop mechanistic multiscale models that integrate system behavior from genomes to landscapes

and subsurface to troposphere. However, developing meaningful models requires more mechanistic information about plant and microbial biology—the organisms, populations, communities, and interactions among and between organisms. This is especially true of the root-rhizosphere phenome, which is key to plant soil interactions, resource efficiency, and thus the sustainability of biofuel agroecosystems.

Emerging research challenges and opportunities in sustainability that can be addressed through the Genomic Science program include:

- Understanding the molecular and physiological mechanisms that control crop vigor, resource use efficiency, and resilience to increase biomass productivity under changing and occasional suboptimal conditions.
- Characterizing the extent to which microbes contribute to plant performance, stress tolerance, and adaptation to changing environmental conditions, and maximizing positive plant-microbe interactions in cropping systems by developing cultivars that take advantage of these beneficial interactions.
- Using genomics knowledge of soil microbial communities to predict the response of key biogeochemical processes such as carbon stabilization, denitrification and nitrous oxide fluxes, methane oxidation, and leaching losses to episodic environmental events such as freeze-thaw cycles, prolonged drought, and rainfall.
- Manipulating specific feedstock plant traits to enhance soil carbon sequestration via changes to the quantity, quality, and location of belowground carbon inputs, and promote nutrient and soil conservation through changes in rooting patterns and architecture.
- Developing process-based, multiscale models that take into account key variables such as genomic properties of biofuel crops, rhizosphere community functions, soil carbon stabilization, nutrient cycling, and greenhouse gas production and that accurately predict plant performance and ecosystem processes under changing environments.

Gaining a fundamental understanding of the molecular and physiological plant processes underlying plant resilience and adaptation to change, as well as the plant-microbe interactions that influence these traits, will enable better optimization of bioenergy crops to marginal landscapes and better predictions of bioenergy crop production across agricultural regions.

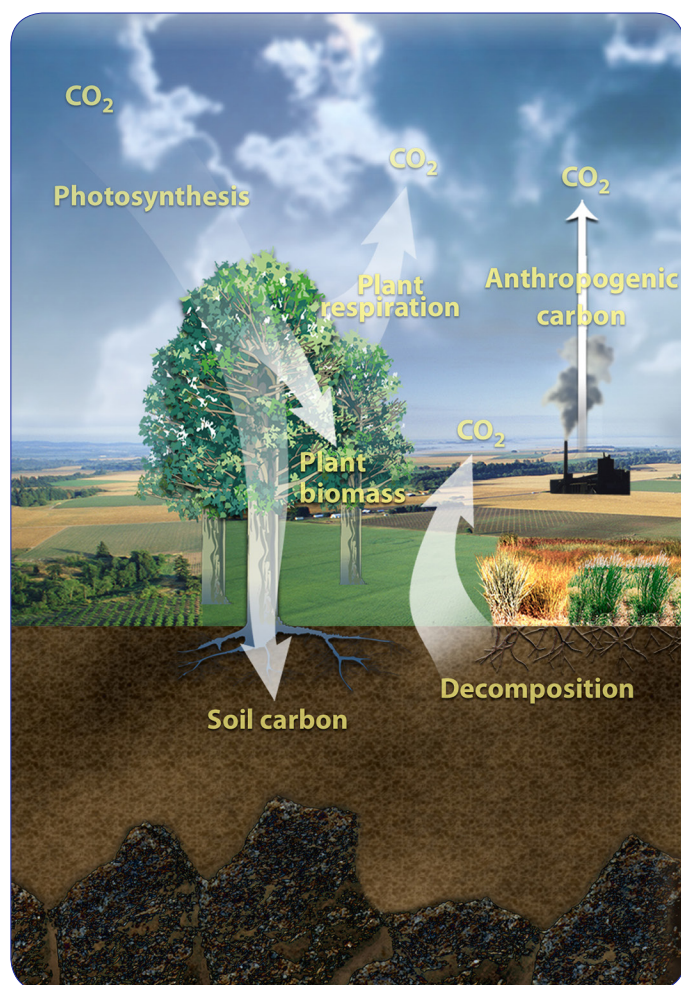


Fig. 5. Processes of the Terrestrial Carbon Cycle.

More information about Genomic Science program objectives for sustainability research can be found at genomic-science.energy.gov/sustainability/.

Systems Biology to Advance Understanding of Carbon Cycling and Other Biogeochemical Processes

Rationale

The global carbon cycle plays a central role in regulating atmospheric carbon dioxide levels and Earth's climate, but knowledge of the biological processes operating at the most foundational level of the carbon cycle remains limited. These processes are intimately linked to higher-scale biogeochemical processes and form key linkages between global carbon, nitrogen, and water cycles. Even minor changes in the rate and magnitude of biological carbon

cycling can have immense impacts on whether ecosystems will capture, store, or release carbon. Developing a more sophisticated and quantitative understanding of molecular-scale processes that drive the carbon cycle represents a major challenge, but this understanding is critical for predicting impacts of global climate change.

The carbon cycle is heavily dependent on microbial communities that decompose or transform organic material in the environment. Massive amounts of organic carbon currently are stored in ecosystems (e.g., the soils of forests, grasslands, and permafrosts; see Fig. 5. Processes of the Terrestrial Carbon Cycle, this page), and microbes are known to play key roles in determining the longevity and stability of this carbon and whether or not it is released into the atmosphere as carbon dioxide or methane, both greenhouse gases, under any given set of environmental variables. However, large uncertainties remain regarding the nature and magnitude of carbon cycle processes performed by environmental microbes; critical knowledge gaps in this area have proven difficult to address given the limited understanding of key groups of organisms and significant technical difficulties in characterizing microbial community functional processes in soil environments. Microbially mediated processes are only minimally represented in most higher-scale carbon cycle models, a factor contributing to uncertainties and limiting predictive capabilities.

Consequently, there is a critical need to advance understanding of the systems biology properties of microbes performing key carbon cycle processes and develop new approaches to link the structural and functional characterization of microbial communities with quantitative measurements of carbon cycle processes.

Systems Biology Approaches

Understanding and predicting processes of the global carbon cycle require bold new research approaches aimed at linking global-scale climate phenomena; biogeochemical processes of ecosystems; and functional activities encoded in the genomes of microbes, plants, and biological communities. This goal presents a formidable challenge, but emerging systems biology research approaches provide new opportunities to bridge the knowledge gap between molecular- and global-scale phenomena. Systems-level research emphasizes studies on the underlying principles of intact, complex systems and facilitates scaling of concepts and data

across multiple levels of biological organization. Applying this approach to the global carbon cycle will require multifaceted, but highly integrated research that incorporates experimentation on model organisms and systems, collection of observational data on communities and ecosystems, and mechanistic modeling of processes ranging from metabolic to global scales.

As described previously, the Genomic Science program approach to systems biology—coupling modeling and simulation with experiment and theory—aims to define the organizing principles that control the functional capabilities of organisms. Researchers have made significant strides in extending these approaches beyond well-characterized model organisms such as *Escherichia coli* and applying them to poorly understood microbes that only recently may have been brought into cultivation from environments of interest. Recent efforts have seen a rapid acceleration in researchers' ability to characterize newly isolated microbes, develop genetic tools for experimental manipulation, and build predictive models of metabolic and regulatory processes. Applying these approaches to key carbon cycling microbes such as methane-consuming bacteria, methane-producing archaea, and syntrophic consortia of organisms responsible for processes that cannot be performed by independent microbes has yielded surprising discoveries of previously unknown capabilities that have important implications for understanding microbes' roles in environmental processes.

To advance this understanding, researchers must move beyond single-organism studies and extend systems biology approaches to more complex microbial communities, the influence of changing environmental variables on functional attributes, and relationships with associated plant communities. Increasingly sophisticated approaches for metagenomics, metatranscriptomics, and metaproteomics (as well as computational tools for associated data analysis and integration) offer new methods for investigating the metabolisms and lifestyles of microbial communities, including uncultivated organisms from environmental samples. Continued development of “meta-omic” research techniques, especially when coupled with approaches permitting analysis and visualization of molecular-scale functional processes, not only will help capture the functional potential encoded in genomes, but also will enable new approaches for qualitatively and quantitatively measuring active processes in the environment. The knowledge gained via these approaches will expand understanding

of molecular-scale carbon cycling processes and their sensitivities to shifting environmental variables, and that knowledge (and, where appropriate, derived data) can be scaled to help refine understanding of processes at higher levels of organization.

The response of ecosystems to environmental change depends on the collective responses of many types of organisms whose functions are encoded in genomes. Although linking directly from genomes to global phenomena is not necessarily practical for terrestrial systems with current approaches, many connections at intermediate scales are viable with integrated application of new systems biology approaches and powerful analytical technologies and modeling techniques aimed at linking physiological and ecosystem-level process understanding. By adapting genome-enabled techniques to the investigation of microbial systems (either via *in situ* observations at field sites or by using micro- or mesocosms for manipulative experiments), this research will pave the way to more quantitative, community-scale systems biology approaches that can inform larger-scale biogeochemical models of microbial processes in the environment.

Specific knowledge gaps in carbon cycling processes that are being addressed by the Genomic Science program include:

- Systems biology studies on microbes, microbial consortia, and microbe-plant interactions involved in large-scale terrestrial carbon cycling processes, with particular emphasis on moving organisms that have been determined to play key roles to experimentally tractable model organism status.
- Determination of the role of microbial communities as key points of integration between major biogeochemical cycles (e.g., carbon, nitrogen, sulfur, and phosphorus) and how these interconnections impact the rate and magnitude of carbon cycle processes in changing ecosystems.
- Further development of environmental omics approaches (e.g., metagenomics, metatranscriptomics, and metaproteomics) and associated data integration tools to investigate *in situ* microbial community functional processes involved in carbon cycling.
- Development of omics-enabled techniques for imaging and analysis of microbial community structure and function in terrestrial environments, with particular emphasis on tools permitting high-resolution and

quantitative measurements in technically challenging settings such as soils, sediments, and key interfacial environments (i.e., decaying organic material, mineral aggregates, and plant roots).

Gaining a mechanistic understanding of the impact microbial communities have on carbon cycling enables more accurate predictions of the feedbacks between the terrestrial environment and the atmosphere under a variety of changing climate scenarios.

In addition to their impact on carbon cycling, microorganisms can profoundly affect the biogeochemical characteristics of soils and sediments, thereby influencing the mobility and fate of materials in the environment. Although numerous processes influence the transport of metals, contaminants, and nutrients, microbiological activity is arguably the least-understood component of these processes. Microbial communities, through their interactions with each other and with geologic materials, play a role in modifying the local geochemical environment, thereby impacting the chemical form and mobility of materials in subsurface and surface environments, particularly at key intersections between the two (thus contributing to broader biogeochemical cycles).

A key to successfully understanding these systems will lie in establishing the link between biological processes performed by microbial communities and site-specific geochemistry. Many systems biology-enabled techniques and approaches described in this section and the section on sustainability (see p. 9) also are applicable to this DOE mission challenge for the Genomic Science program, and were in many cases initially developed for studies of subsurface microbial communities.

Improving the mechanistic understanding of key biogeochemical processes mediated by microbial communities facilitates the development of more predictive field-scale models of environmental processes. Microbial processes often are abstracted as static descriptions of individual enzymatic processes, yet the influence of community composition, growth, stress, and nutrient limitation (among many others) can greatly affect activity across spatial and temporal scales. New systems biology-enabled understanding offers the ability to describe how individual microbes and microbial communities function in response to changing environmental parameters and vice versa. Furthermore, the integration of systems biology and modeling approaches that link omics-derived data on microbial community

structure and function with data from biogeochemical and environmental measurements will provide new mechanistic ways to more accurately describe and predict the dynamic interplay between microbes and their environment.

The Genomic Science program will continue to facilitate detailed discovery and investigation of microbes and microbial ecosystems that play roles in the fate and transport of metals, contaminants, and nutrients in subsurface environments. These studies will expand knowledge about structure, function, metabolic activity, and the dynamic nature of microbial communities and their interaction with the geochemical environment. Research on microbial community structure and function in these environments also provides opportunities to further develop community-scale systems biology tools linked to biogeochemical process understanding and pose more foundational questions relating to ecological theories and their applicability to microbial communities.

More information on Genomic Science program objectives in carbon cycle and biogeochemical research can be found at genomicscience.energy.gov/carboncycle/.

Cross-Cutting Computational Biology, Bioinformatics, and Predictive Modeling

The scientific objectives of the Genomic Science program require a highly coordinated application of expertise that transcends traditional disciplinary boundaries. As such, one of the program's most challenging but critical goals is the creation of robust computational frameworks for data integration, analysis, and sharing that can accommodate the wide variety of heterogeneous data streams being generated across the Genomic Science community. These frameworks include not only the various types of omics data (as well as meta-omics variations) discussed earlier in this report, but also data derived from nonomics-based analytical technologies for quantitative physiological analysis, physicochemical measurements of environmental factors, and a vast array of other experimental data types.

Data-specific needs for Genomic Science program research often revolve around tracking high-throughput experimental and contextual environmental data; developing tools for capturing and archiving large and complex datasets; and generating innovative new approaches for analysis, distillation, and integration of systems biology

data. Tracking the data requires a Laboratory Information Management System (LIMS) appropriate for specific project needs to monitor experimental cycles, track samples and workflows, and collect internally compatible data from varying instrument types. Data capture and storage present considerable challenges, and the volume and complexity of data generated by systems biology research often require new technologies and bioinformatics approaches permitting rapid data storage, retrieval, and transfer at very large scales. The generation of raw data only begins the cycle of scientific inquiry. Improved data-distillation strategies for filtering out noise and compressing noncritical information, as well as identifying biologically meaningful data subsets, are critical for enabling subsequent cycles of analysis, integration, and modeling.

The process of modeling and simulation attempts to build a more integrated understanding of the dynamic nature of biological systems and enable scientists to test their knowledge via computerized “virtual experiments.” Creating models that predict biosystem response to untested conditions requires continued emphasis on quantitative details such as kinetic constants, enzyme activities, and dynamic metabolic measurements underlying functional biological processes. Continuing to build on well-developed model organisms such as *Escherichia coli*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana* is important, as is extending this area of research to develop predictive models of biological function in a broader class of organisms. Moving beyond the level of individual organisms, new mathematical and machine-learning methods are needed to address biological variables at the community scale and understand evolving

interactions with external signals from the environment. As more powerful resources for high-performance computing become available, the amount of biological data produced by high-throughput experimental approaches grows at an even faster pace. Although this data continues to yield insights into and improve the quantitative understanding of biological systems, incorporating detailed molecular, biochemical, physiological, and structural information into biological models and simulations remains a major challenge.

Given the data-intensive nature of Genomic Science research, all supported projects are required to generate data management and integration plans that emphasize an iterative approach to data analysis and lead to a predictive understanding of the biological system(s) under investigation. Developing these types of plans serves not only the objectives of the individual project, but also facilitates the collaborative sharing of resulting data across the broader community via mechanisms such as the DOE Systems Biology Knowledgebase (see sidebar, DOE Systems Biology Knowledgebase, p. 16). The long-term success of the Genomic Science program, and systems biology in general, depends on achieving high levels of data and information integration and sharing. BER has established an information and data sharing policy requiring public accessibility to all publishable information (see sidebar, Genomic Science Information and Data Sharing Policy, this page).

More information on Genomic Science program efforts on data and information integration and sharing can be found at genomicscience.energy.gov/compbio/.

Genomic Science Information and Data Sharing Policy

The Department of Energy’s (DOE) Office of Biological and Environmental Research requires that all publishable data, metadata, and software resulting from research funded by the Genomic Science program must conform to community-recognized standard formats when they exist, be clearly attributable, and be deposited within a community-recognized public database(s) appropriate for the research (genomicscience.energy.gov/datasharing/). Publication-related data that are consistent with the data and analysis models of the DOE Systems Biology Knowledgebase (KBase) should be accessible

to KBase and could be integrated into KBase. All digitally accessible data obtained as a result of research funded by the Genomic Science program must be kept in an archive maintained by the principal investigator for the duration of the funded project. Any publications resulting from the use of shared data must accurately acknowledge the original source or provider of the attributable data. The publication of information resulting from research funded by the Genomic Science program must be consistent with the intellectual property provisions of the contract under which the publishable information was produced.

DOE Systems Biology Knowledgebase

kbase.us

Community-Driven Cyberinfrastructure for Sharing and Integrating Data and Analytical Tools to Accelerate Predictive Biology

In 2011, the Genomic Science program launched the Department of Energy (DOE) Systems Biology Knowledgebase (KBase; kbase.us), an open bioinformatics platform for predictive systems biology designed to accelerate understanding of microbes, microbial communities, and plants relevant to DOE's missions (see Fig. 6. DOE Systems Biology Knowledgebase, p. 17). As a community resource, KBase's purpose is to integrate a wide spectrum of genomics and systems biology data, models, and bioinformatics tools to ultimately predict and design biological function. KBase allows researchers to collaboratively generate, test, and share new hypotheses about gene and protein functions; perform large-scale analyses on a scalable computing infrastructure; model interactions between relevant organisms; and propose new experiments to further refine the models.

KBase allows users to upload their own data, access public data, and securely share and publish their workflows and conclusions, providing transparency and reproducibility for the complex, iterative, multidisciplinary research to address biological challenges faced by DOE in energy and the environment.

KBase currently is supported by a computing infrastructure based on OpenStack cloud system software distributed across core sites at four DOE national laboratories and leverages DOE's Energy Sciences Network (ESNet) for data transmission. Distinctive features include:

- One-stop destination for integrative analysis of data from plants, microbes, and their communities.
- Biological data from existing databases linked by KBase infrastructure and integrated into a comprehensive data model to accelerate large-scale data analyses and facilitate comparative studies of complex datasets.
- Graphical user interface designed to facilitate communication and collaboration among researchers by enabling sharing of analytical workflows and procedural notes, visualizations, custom scripts, and results.
- Tools enabling users to access KBase resources from their own systems or develop new tools for broader use within the KBase infrastructure.

To achieve high-quality predictions, it is imperative that biological data and associated metadata also be high quality. KBase currently supports more than 900 data types (kbase.us/data/data-types/); as this number grows, it also will include user-specific data types, once they pass quality and metadata standards.

The large scale and complexity of Genomic Science program approaches generate enormous amounts of data and information at a rate that far outpaces current analytical capacity. Moreover, Genomic Science program research requires large, multidisciplinary teams that will benefit from well-coordinated sharing and use of the massive datasets that they produce. KBase will facilitate collaboration and data sharing not only within such multidisciplinary consortia, but also between them and individual research efforts that will gain access to dramatically more data and robust analytical and modeling tools that otherwise may not be available to smaller, individual projects. As such, KBase addresses several critical needs and should significantly accelerate the pace of discovery and predictive understanding.

A vast and diverse range of biological systems and their environments is relevant to DOE missions. Understanding the environmental conditions that influence an organism's functional properties requires metadata with contextual information such as the microenvironment immediately surrounding that organism. Availability of a common resource such as KBase for collectively gathering data and experimental results will facilitate community-wide efforts to establish the guidelines and standards needed to adequately capture environmental metadata.

Given that all life forms, at the molecular level, are based on similar sets of fundamental processes and principles, the understanding of one biological system under study can be advanced by integrative analyses involving knowledge from other organisms. KBase provides an open infrastructure for mining, comparing, and interconnecting large biological and environmental datasets, making possible the comprehensive knowledge needed to predict how the complex interplay between genomes and environments controls the behavior of biological systems.

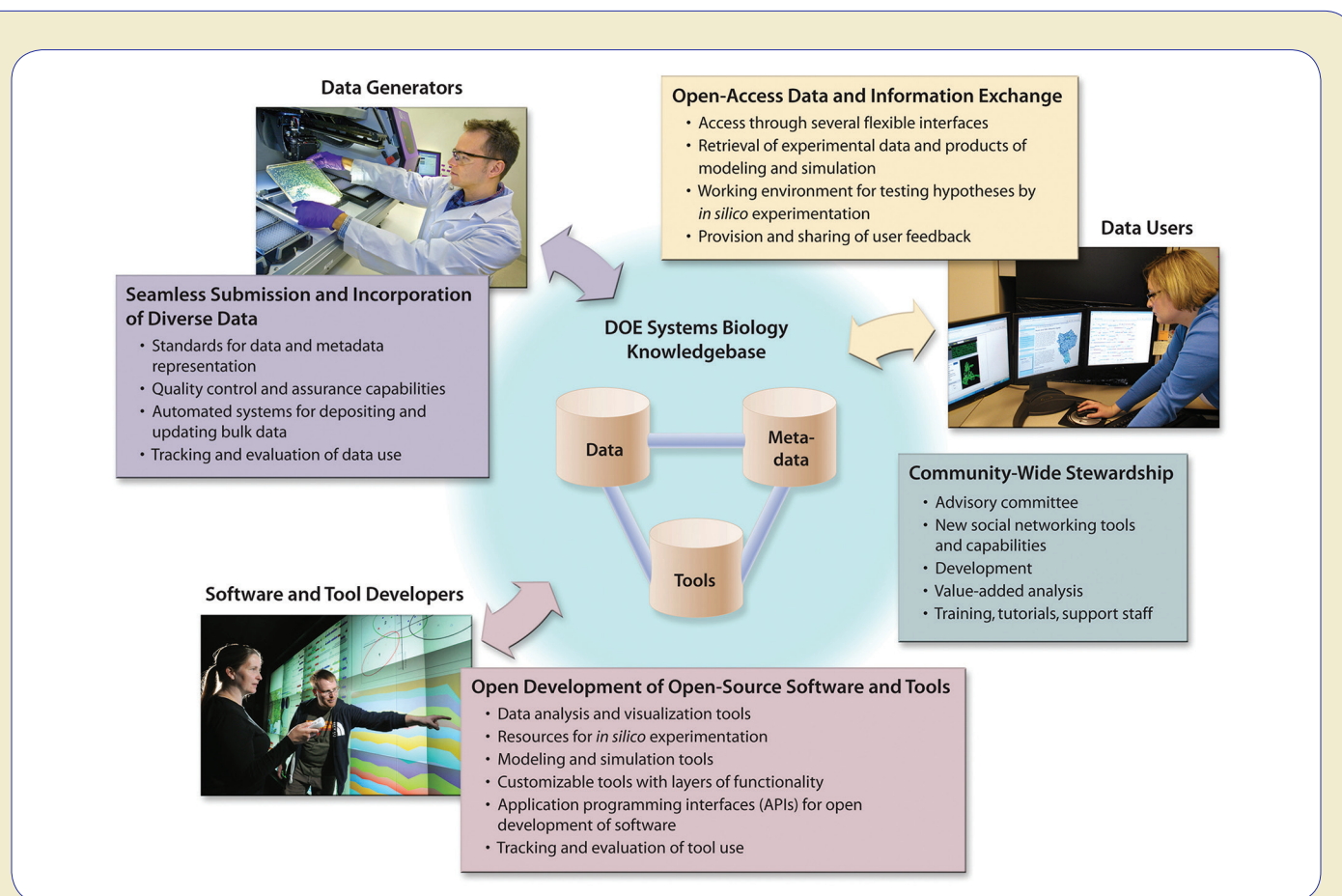


Fig. 6. DOE Systems Biology Knowledgebase.

Programmatic Context for DOE Genomic Science

DOE Office of Science

The Department of Energy's (DOE) Office of Science manages fundamental research programs in basic energy sciences, high-energy physics, fusion, biological and environmental sciences, and computational science, supporting unique and vital parts of U.S. research in climate change, geophysics, genomics, and life sciences.

The Office of Science also manages 10 world-class national laboratories with unmatched capabilities for solving complex interdisciplinary scientific problems and oversees the construction and operation of some of the nation's most advanced scientific user facilities, located at national laboratories and universities. These include particle and nuclear physics accelerators, synchrotron light sources, neutron scattering facilities, supercomputers and high-speed computer networks, genome and protein sequencing facilities, and advanced resources in imaging and analysis for biological and environmental systems.

DOE Office of Biological and Environmental Research

Office of Biological and Environmental Research (BER) programs within the DOE Office of Science advance world-class research to understand complex biological and environmental systems and provide scientific user facilities (see sidebar, BER National User Facilities, p. 20) to support DOE missions in scientific discovery and innovation, energy security, and environmental responsibility. BER's interdisciplinary research programs engage scientists from national laboratories, universities, and the private sector in leveraging diverse scientific insights by coupling theory, observations, experiments, models, and simulations.

The programs are managed within two divisions: the Biological Systems Science Division (BSSD) and Climate and Environmental Sciences Division (CESD). BSSD seeks to understand how genomic information is translated to functional capabilities, enabling more confident redesign of microbes and plants for sustainable biofuel production, improved carbon storage, and a better understanding of biological transformation of materials such as nutrients and contaminants in the environment. CESD advances understanding of the roles of Earth's biogeochemical systems (i.e., the atmosphere, land, oceans, sea ice, and sub-surface) in determining climate to facilitate prediction of climate decades or centuries into the future—information needed to plan for energy and resource needs.

Genomic Science Program Management Principles

BER's Genomic Science research portfolio includes discovery-driven science, hypothesis-driven science, technology development, and foundational research. Management of the Genomic Science program is guided by the following principles:

- Maintain a strategically managed research portfolio responsive to both existing and emerging national priorities and mission needs.
- Engage collaboratively with the broader scientific community to identify key knowledge gaps and inform research priorities for DOE mission science via workshops and other strategic planning activities.
- Issue funding solicitations that emphasize innovative, high-risk and high-reward systems biology research.
- Select research projects based on mission relevance, scientific merit, and peer review.
- Support research conducted by individual investigators, collaborative project teams, and integrated multidisciplinary research centers, drawing on the distinctive strengths of scientists at DOE national laboratories, academic institutions, and industrial partners.
- Encourage communication across the scientific community through the annual Genomic Science program principal investigator meeting, workshops focused on specific issues or research topics, symposia, and exhibits at national meetings.
- Leverage capabilities and resources across BER programs and DOE scientific user facilities.
- Foster an atmosphere of open information exchange through the development of community resources that facilitate the integration and sharing of data resulting from funded research efforts.
- Contribute to developing a scientific workforce with the multidisciplinary expertise necessary to address DOE mission challenges.
- Coordinate with other DOE programs and other federal agencies to leverage scientific research investments.

BER National User Facilities

The Department of Energy's (DOE) Office of Science creates, maintains, and operates state-of-the-art national user facilities that are key to continued U.S. leadership in physical and biological research. Transformational scientific discoveries are resulting from biological research at the Office of Science's advanced scientific computing, synchrotron, advanced light source, and neutron facilities. These world-leading facilities are enabling scientists to track biomolecular processes in real time and image biological materials at atomic resolution. User facilities focused on genome sequencing and environmental molecular research are supported by the Office of Biological and Environmental Research (BER) within the Office of Science.

DOE Structural Biology User Facilities

genomicscience.energy.gov/userfacilities/structuralbio.shtml

BER supports end stations for structural biology at DOE light sources and neutron facilities to provide reliable access to these major user facilities for research in the life sciences. This activity is carried out in coordination with the DOE Office of Basic Energy Sciences, National Institutes of Health, National Science Foundation, and other public and private organizations. BER-supported resources are:

- Macromolecular Crystallography Research Resource at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory
- Structural Molecular Biology Center at the Stanford Synchrotron Radiation Lightsource (SSRL) and the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory
- National Center for X-Ray Tomography, Berkeley Synchrotron Infrared Structural Biology Program, Advanced Biological and Environmental X-Ray Spectroscopy, and Structurally Integrated Biology for the Life Sciences Beamline, all located at the Advanced Light Source (ALS), Lawrence Berkeley National Laboratory
- Structural Biology Center at the Advanced Photon Source (APS), Argonne National Laboratory
- Center for Structural Molecular Biology at the High Flux Isotope Reactor (HFIR) and Spallation Neutron Source (SNS), Oak Ridge National Laboratory

Each of these BER-supported resources is heavily utilized, and all are providing valuable insights for a variety of DOE-supported research activities related to the Genomic Science program. BER involvement at these user facilities enables user access at new Office of Science facilities (e.g., LCLS, SNS, and NSLS-II) and future facilities by funding new beam lines and associated laboratory and computational capabilities to enable utilization of these resources by the life sciences community.



Fig. 7. Fourier Transform Infrared Microscopy Instrumentation at the Advanced Light Source. [Image courtesy Lawrence Berkeley National Laboratory]

DOE Joint Genome Institute

jgi.doe.gov

The DOE Joint Genome Institute (JGI) in Walnut Creek, California, is one of the world's largest and most productive public genome-sequencing centers. JGI's sequencing capacity exceeds 70 billion DNA base pairs per year. The majority of this capacity is focused on sequencing plants, microbes, and microbial community metagenomes relevant to DOE energy and environmental missions. New technology development efforts seek to enhance JGI's ability to infer biological meaning from the genome sequences it produces.

DOE JGI's sequencing targets have included a wide range of bacteria and bacterial communities involved in the biogeochemical cycling of carbon and nutrients, detoxification of contaminants, and capture of carbon from the atmosphere. JGI also has sequenced numerous varieties of plants relevant to bioenergy production, as well as microbial communities with degradative abilities for processing plant biomass into advanced biofuels. This growing collection of sequence data will enable discovery of the fundamental principles and details of systems biology needed to develop biobased solutions to critical national needs and DOE missions.

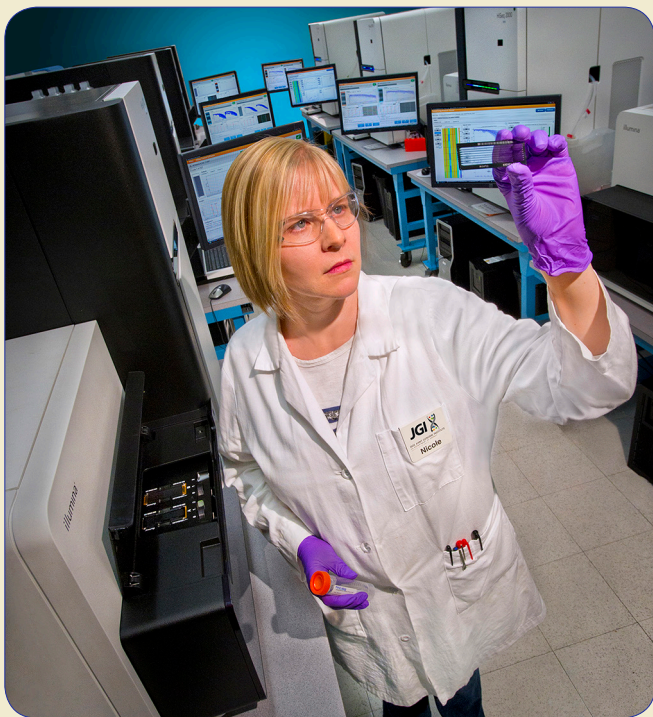


Fig. 8. Illumina HiSeq Sequencing Units and Flow Cells. [Image courtesy DOE JGI]

DOE Environmental Molecular Sciences Laboratory

emsl.pnl.gov

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL), located in Richland, Washington, provides the scientific community with co-located experimental and computational resources to enable fundamental studies of physical, chemical, and biological processes and their interactions to address DOE's energy, environmental, and science missions. With more than 60 leading-edge experimental instruments and a high-performance computer, EMSL offers capabilities for investigating a wide range of issues and processes involved in systems biology studies, including microbial communities and their interfaces with minerals or contaminants.

Relevant capabilities include high-resolution nuclear magnetic resonance (NMR) spectrometers for protein structure studies; high-performance mass spectrometers for proteomics analyses; Mossbauer spectrometers and surface-analysis tools (e.g., a cryotransmission electron microscope and other high-resolution microscopy systems and a secondary ion mass spectrometer) for characterizing or visualizing microbe-mineral or microbe-contaminant interfaces; a flow cell facility for studying particle and microbial transport through groundwater and sediments; and a supercomputer for conducting molecular dynamics studies of protein-protein interactions and protein folding and other biological structure-function research and for simulating subsurface microbial and contaminant transport.



Fig. 9. Time-of-Flight Secondary Ion Mass Spectrometer. [Image courtesy DOE EMSL]

For More Information

Definitions of Terms

DNA synthesis: Process for producing copies of specific DNA molecules, enabling directed exploration of gene and genomic manipulations of biological functions relevant to Department of Energy missions.

Genomics: Study of the entire set of DNA sequences—both coding and noncoding DNA.

Metabolomics: Analysis of the metabolome—the complete set of an organism’s metabolites (small molecular products of cellular processes).

Metagenomics: Study of the collective DNA isolated directly from a community of organisms living in a particular environment.

Omics: Collective term for system-wide biological analyses that include, but are not limited to, genomics, transcriptomics, proteomics, and metabolomics.

Proteomics: Analysis of the proteome—the complete set of proteins expressed by a cell or population of cells.

Systems biology: Integrates data from various omic analyses using computational tools to build predictive models of biological systems.

Transcriptomics: Analysis of the transcriptome—the complete set of RNA molecules present in a cell or population of cells.

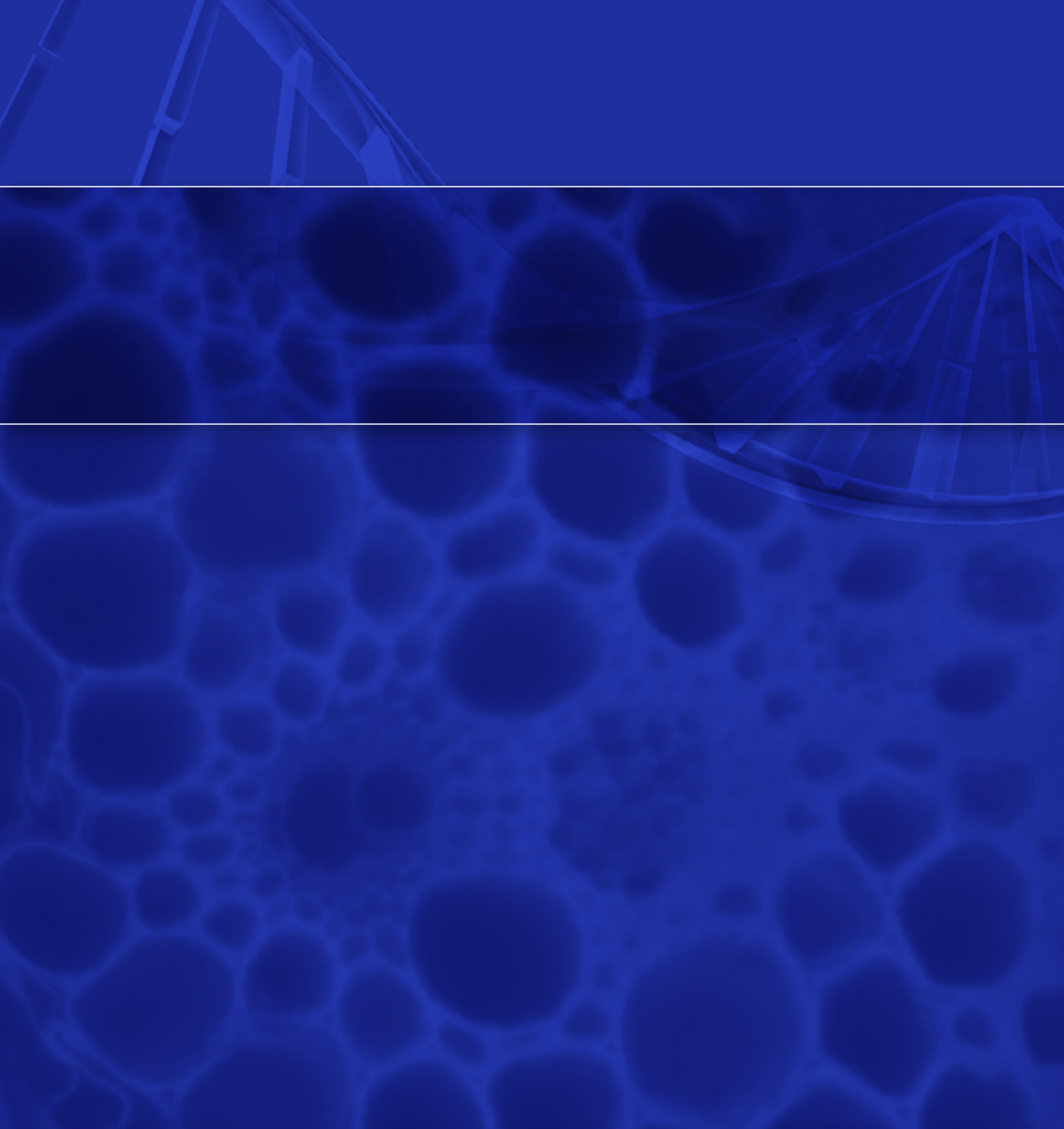
Websites

Genomic Science Program: genomicscience.energy.gov

DOE Office of Biological and Environmental Research: science.energy.gov/ber/

DOE Office of Science: science.energy.gov

U.S. Department of Energy: energy.gov



October 2014